2019

Poonam Dhoot

10399137

5/3/2019

2004 Cars Dataset – Interactive Visualizations



# Introduction

We have used a dataset of different types of cars and trucks. Specifications are given for 427 new vehicles for the year 2004. The attributes include price, fuel efficiency, quantitative measurements relating to the size of the vehicle.

In this dataset, we have 427 observations with 27 variables.

Dataset: 2004 Cars

|  |  |
| --- | --- |
| VehicleName | Model Name |
| TypeOfVehicle | Model Type |
| Company | Brand of Car |
| Country | Origin of Brand |
| Sedan | Yes/No |
| Sports\_Car | Yes/No |
| SUV | Yes/No |
| Wagon | Yes/No |
| Minivan | Yes/No |
| Pickup | Yes/No |
| AWD | All-Wheel Drive? Yes/No |
| RWD | Rear-Wheel Drive?  Yes/No |
| Retail\_Price | Suggested Retail Price, what the manufacturer thinks the vehicle is worth, including adequate profit for the automaker and the dealer |
| Dealer\_Cost | Dealer Cost (or "invoice price"), what the dealership pays the manufacturer (U.S. Dollars) |
| Engine\_Size | Engine Size (liters) |
| Cyl | Number of Cylinders |
| HP | Horsepower |
| City\_MPG | City Miles Per Gallon |
| Hwy\_MPG | Highway Miles Per Gallon |
| Weight | Weight (Pounds) |
| Wheel\_Base | Wheel Base (inches) |
| Len | Length (inches) |
| Width | Width (inches) |
| Lat | Latitude |
| Long | Longitude |
| Dimension | Length x Width |
| Avg\_MPG | Mean of City\_MPG & Hwy\_MPG |

GitHub Link - <https://github.com/pdhoot16/DataViz>

Tools for Visualization

Python, Jupyter - Colab

\*\*Libraries : Pandas, Numpy, Altair (For Interactive Visualizations)\*\*

Contribution

Each team member has contributes and fulfilled the following tasks.

Poonam Dhoot (10399137) --------------------------------------------------------------------

1. Environment Set Setup (Installation Of Packages / Libraries)
2. Data Preprocessing (Data Cleaning , Shaping , Working With Pandas, Altair)
3. Interactive Dashboard With All Interactive Data Visualization
4. Report

Sunmeet Thapar (10506082)-------------------------------------------------------------------

1. Data Preprocessing (Data Cleaning , Shaping)
2. Interactive Dashboard With All Interactive Data Visualization

Ramya Hunasghatta M (10388022)---------------------------------------------------------------

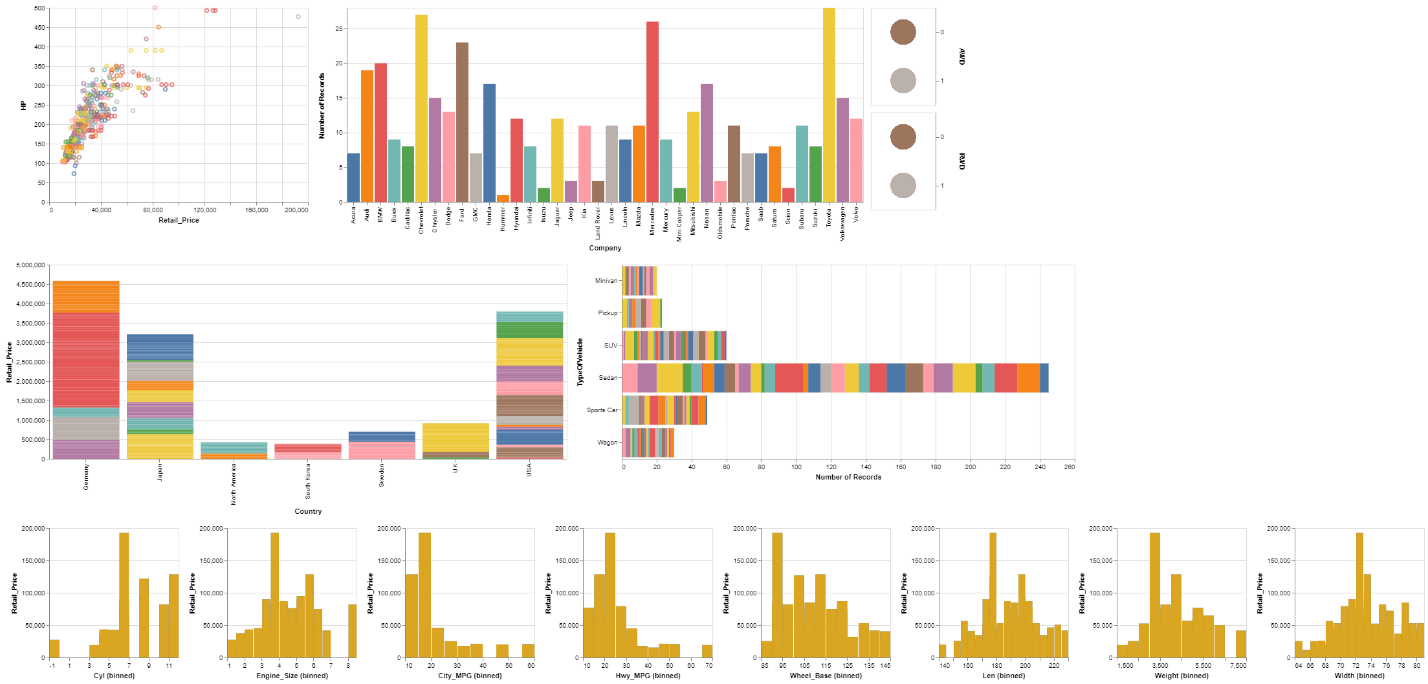
1. Determining Initial Questions For Analysis

Kriti Dhyani (10505759)-----------------------------------------------------------------------

1. Determining Initial Questions For Analysis
2. Getting Insights about the data and which attributes are related to each other

**Dataset Analysis:**

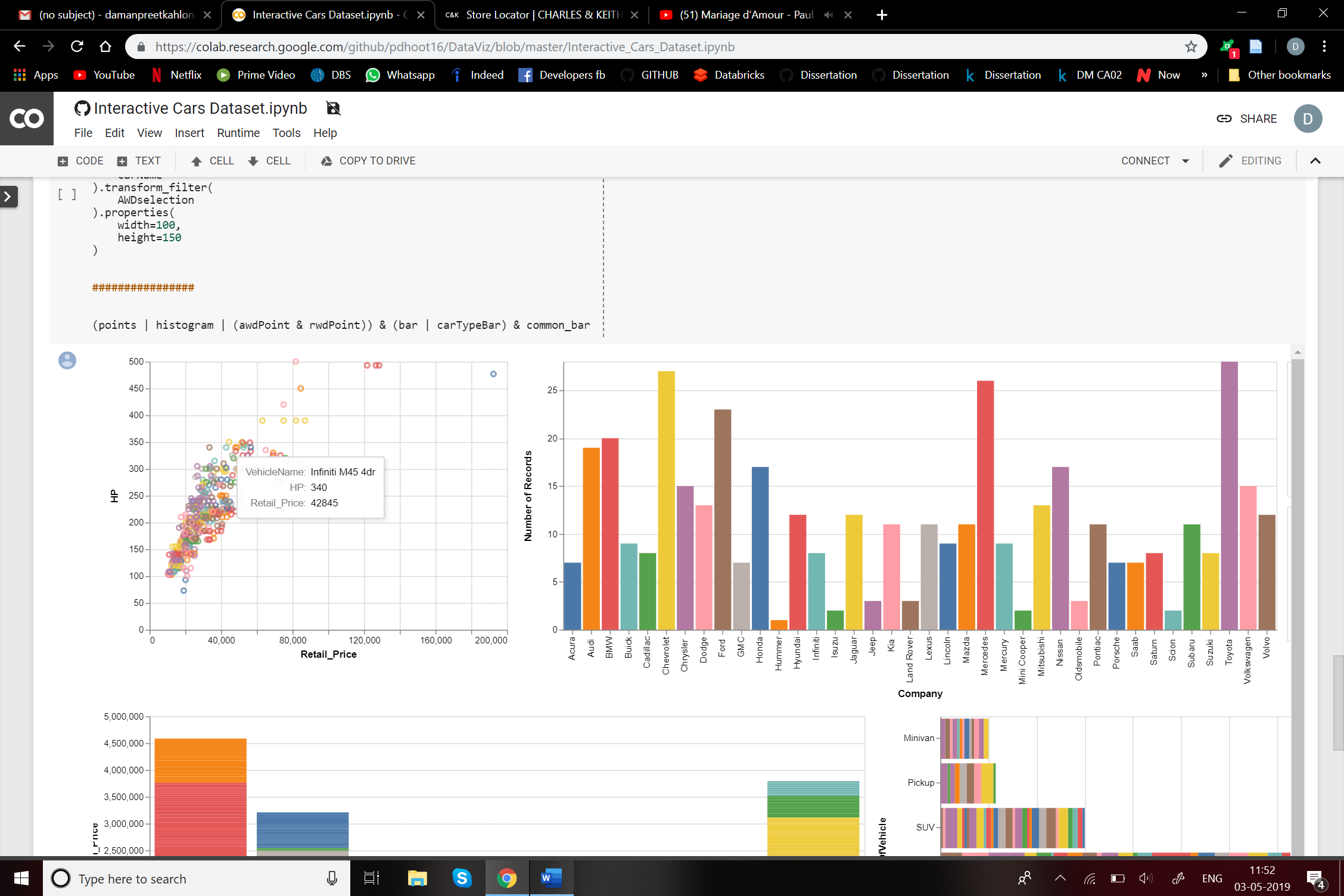
We have created a dashboard in Python that links all our visualizations and analysis.



**To understand, let’s look at the individual aspects of the dashboard**

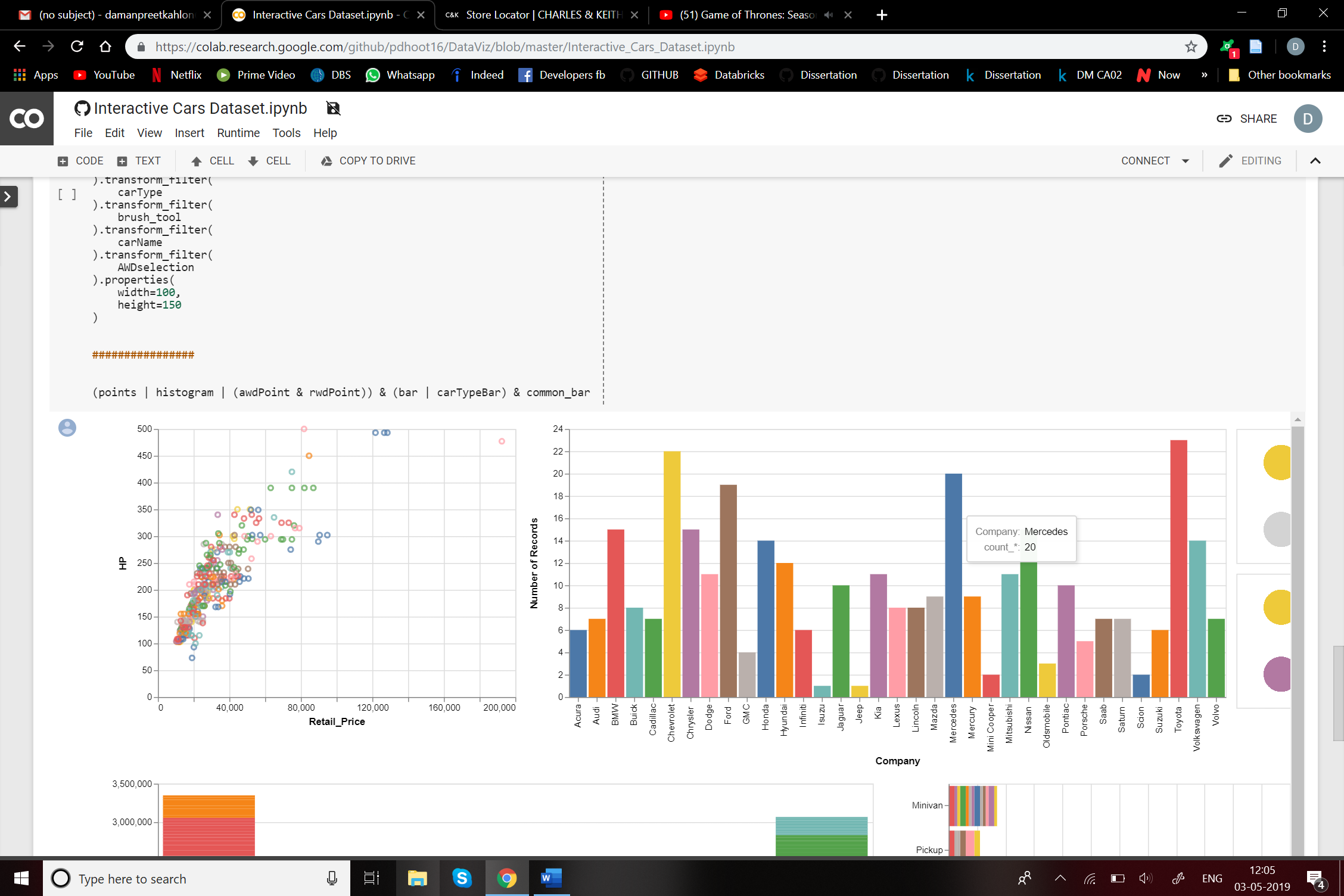
1. **Horse power vs Retail price($):**

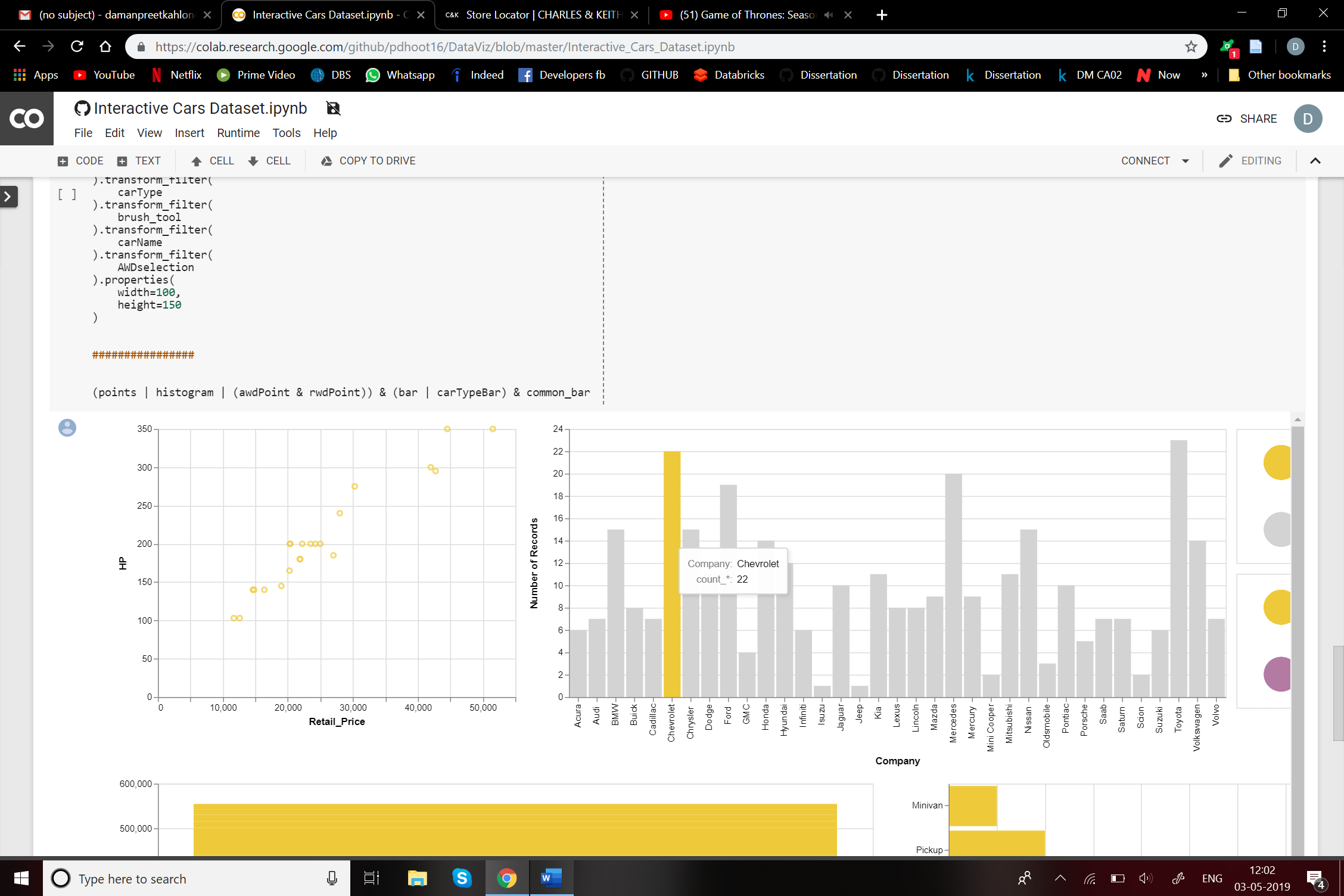
This scatter plots compares the horse power and retail price of each vehicle.



1. **Number of cars of each company:**

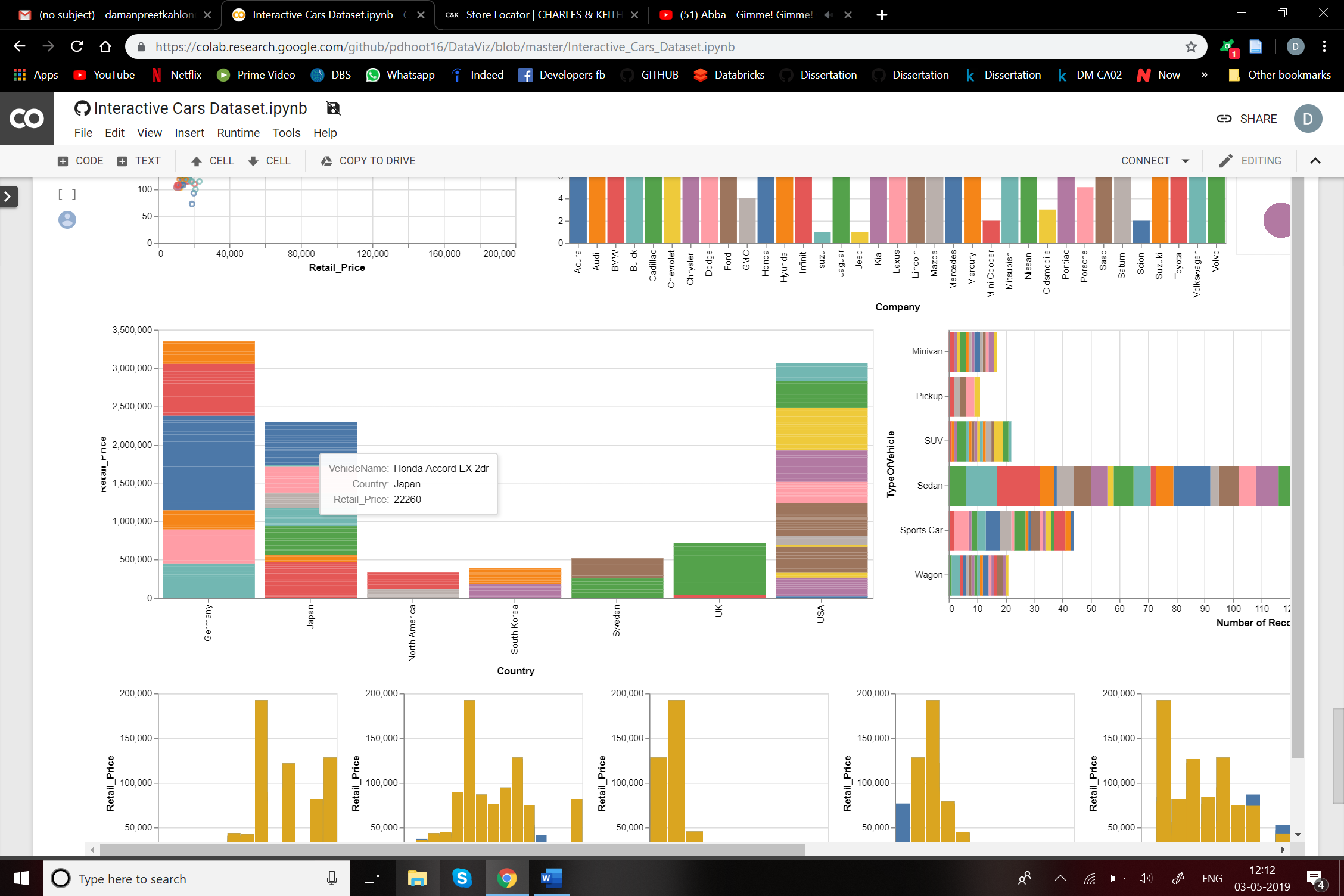
This bar graph gives number of records of different companies. Toyota, Mercedes and Chevrolet have the high number of records. On selecting each record the dashboard highlights the features of the particular company.

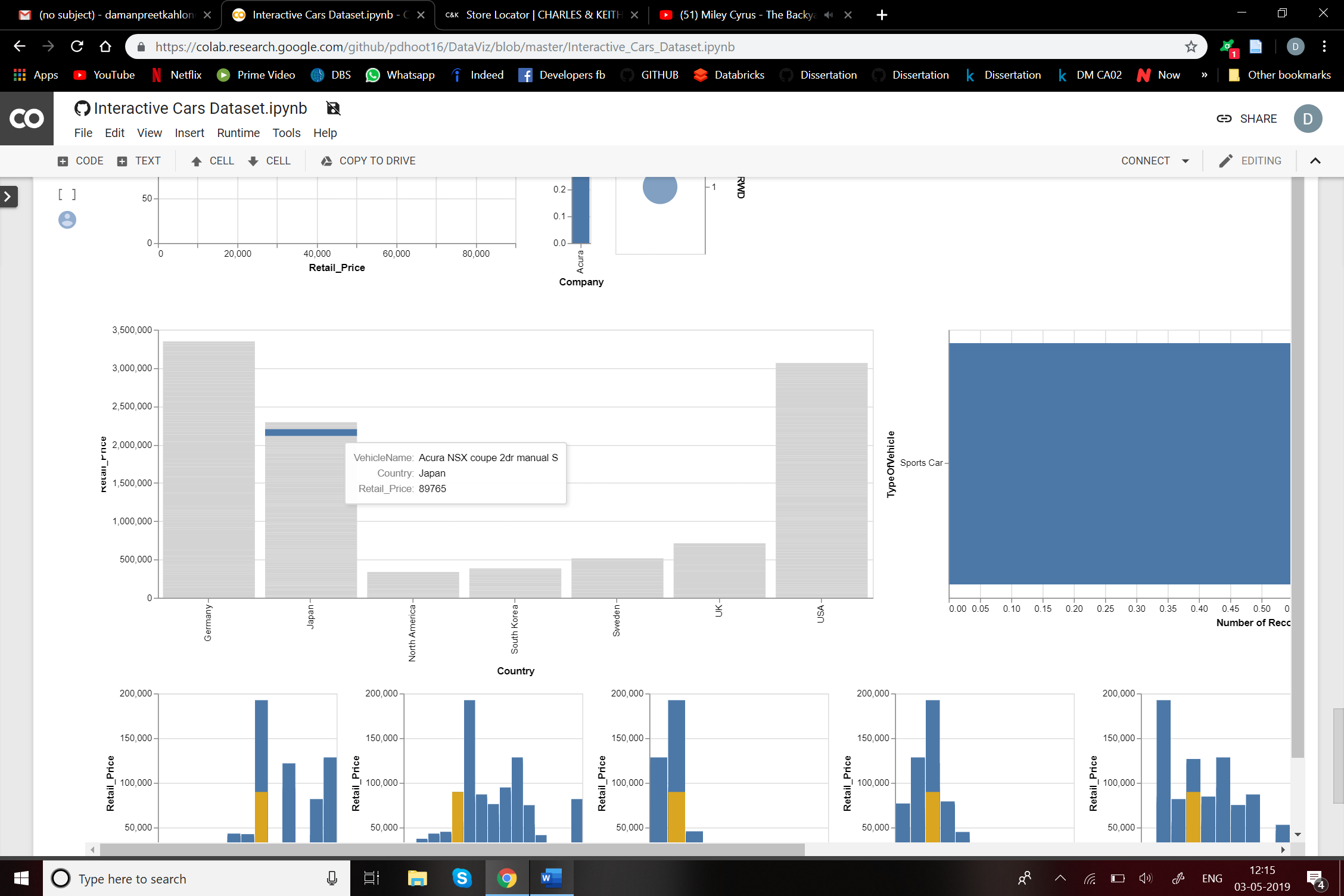




1. **Country wise analysis:**

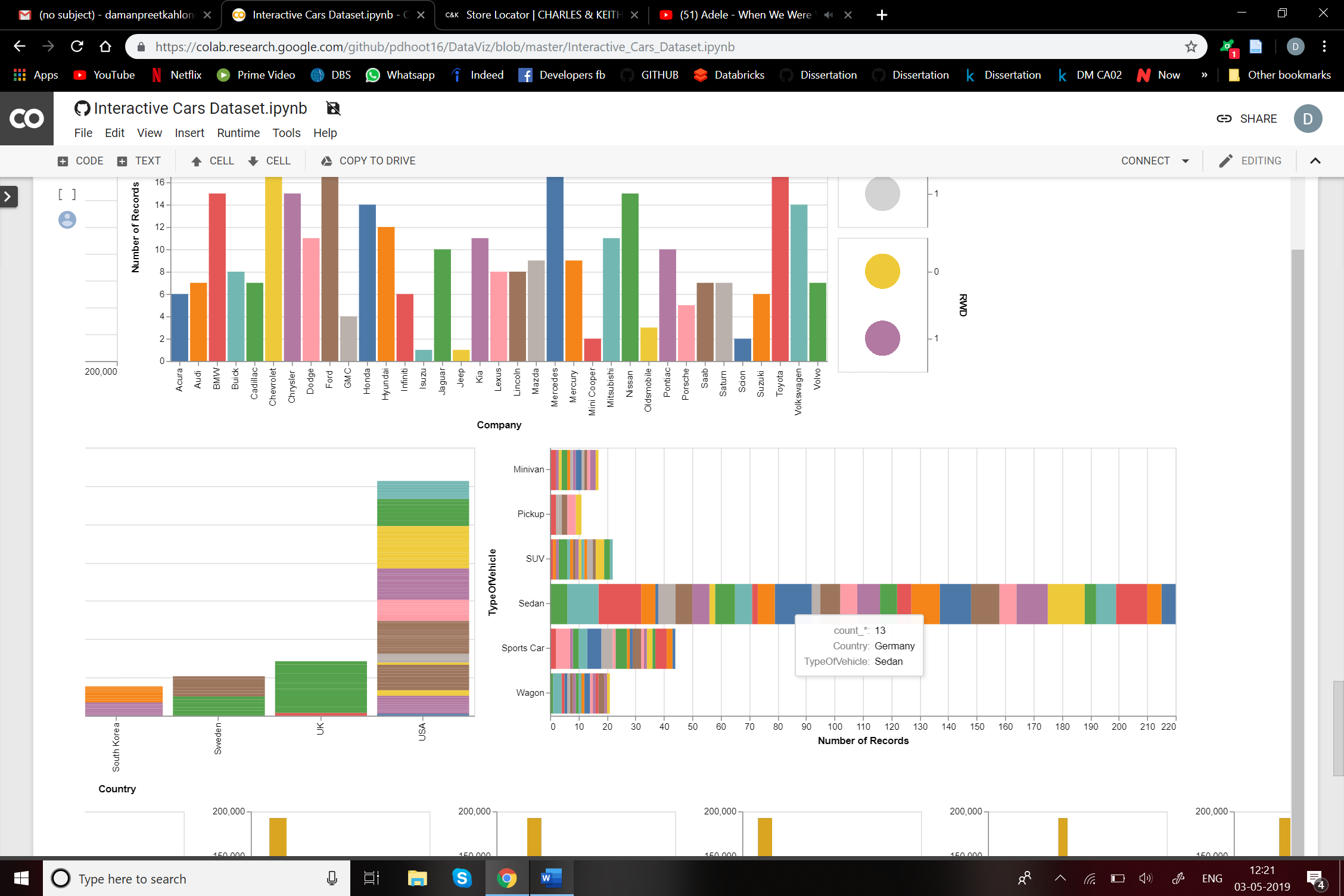
This graph gives the country of the vehicle name along with the retail price. It is known that Germany, USA and Japan are leading manufacturers of cars. These countries manufacture cars for all price ranges. You can find maximum diversity for these countries.

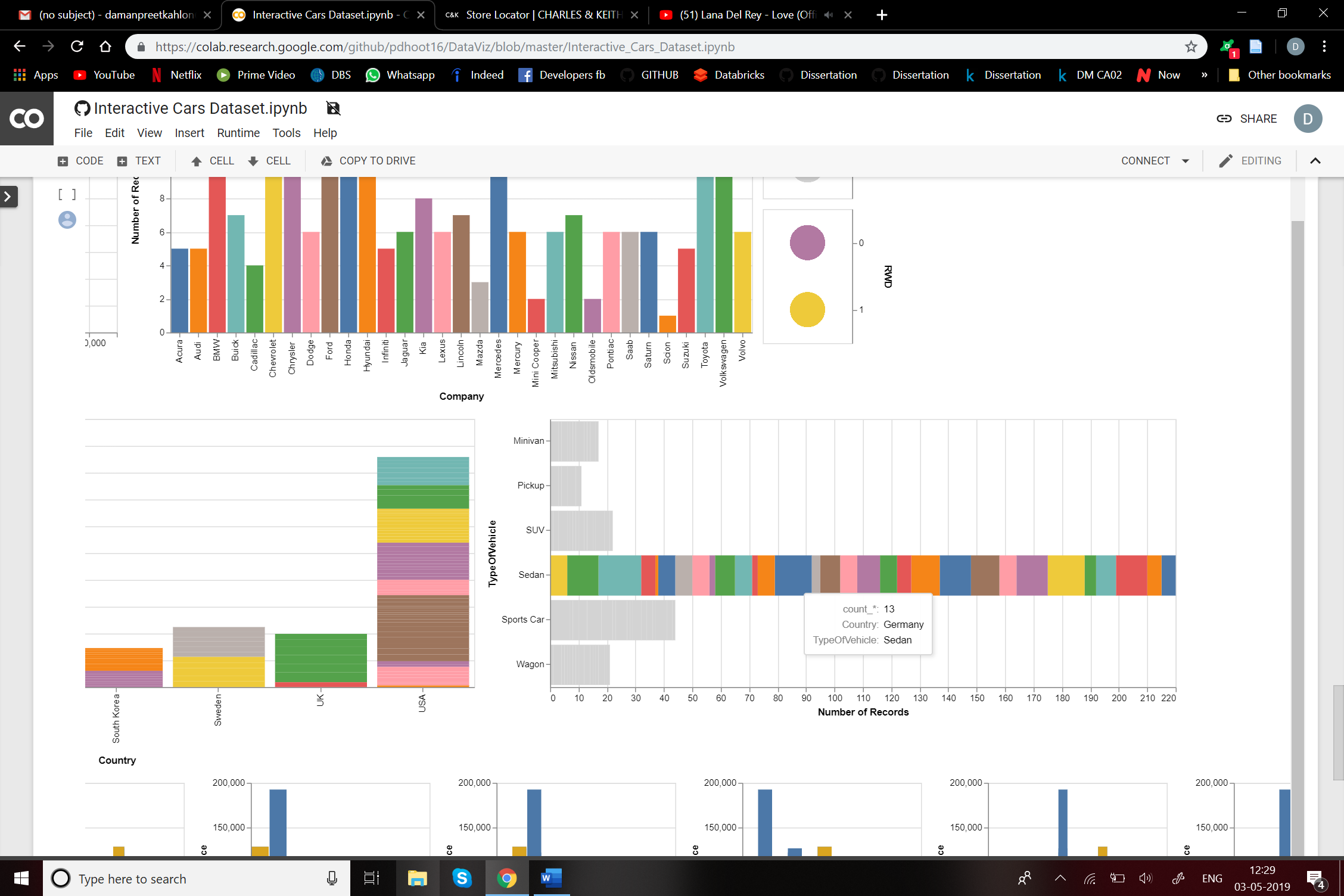




1. **Type of vehicle analysis:**

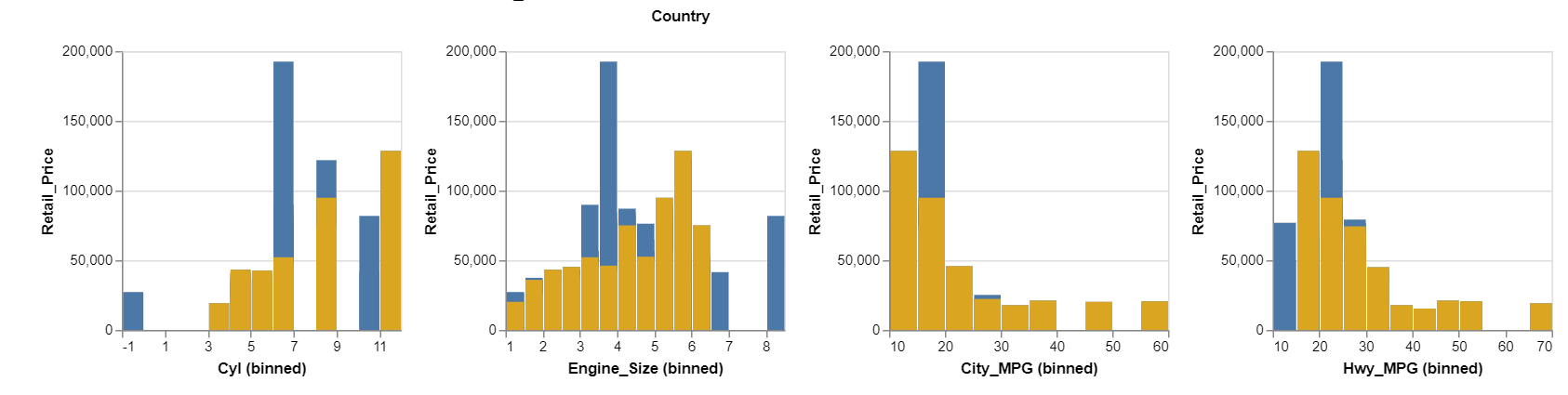
From the dataset, we also get the type of vehicle. They can be classified into Wagon, Sports Car, Sedan, SUV, Pickup and Minivan.

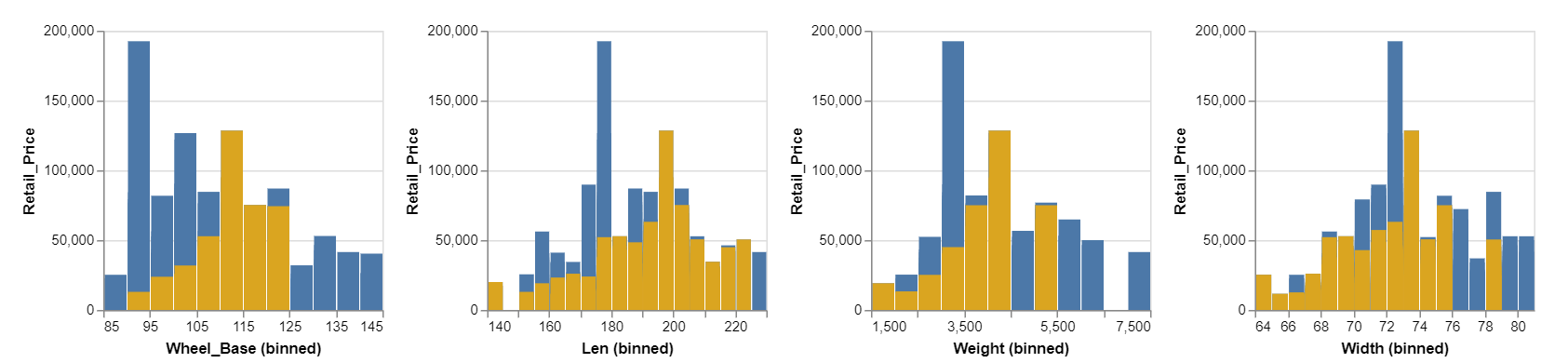




1. **Analysis on the factors that affect the price of the vehicle:**

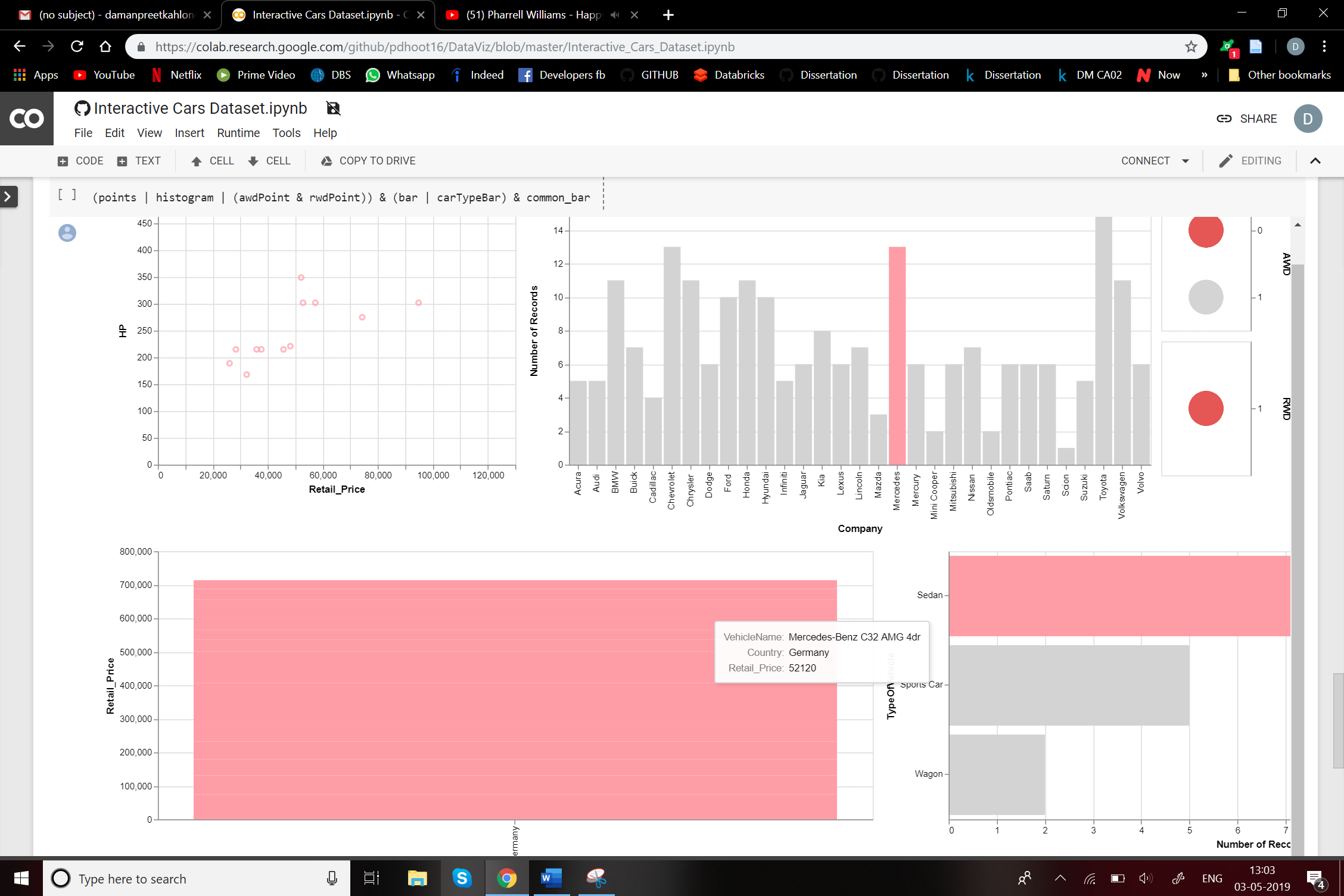
In order to analyse the factors affecting the price of the vehicle, we have plotted Cylinder, Engine Size, City Miles per gallon, Highway Miles per gallon, Wheel base, Len, Weight and Width.



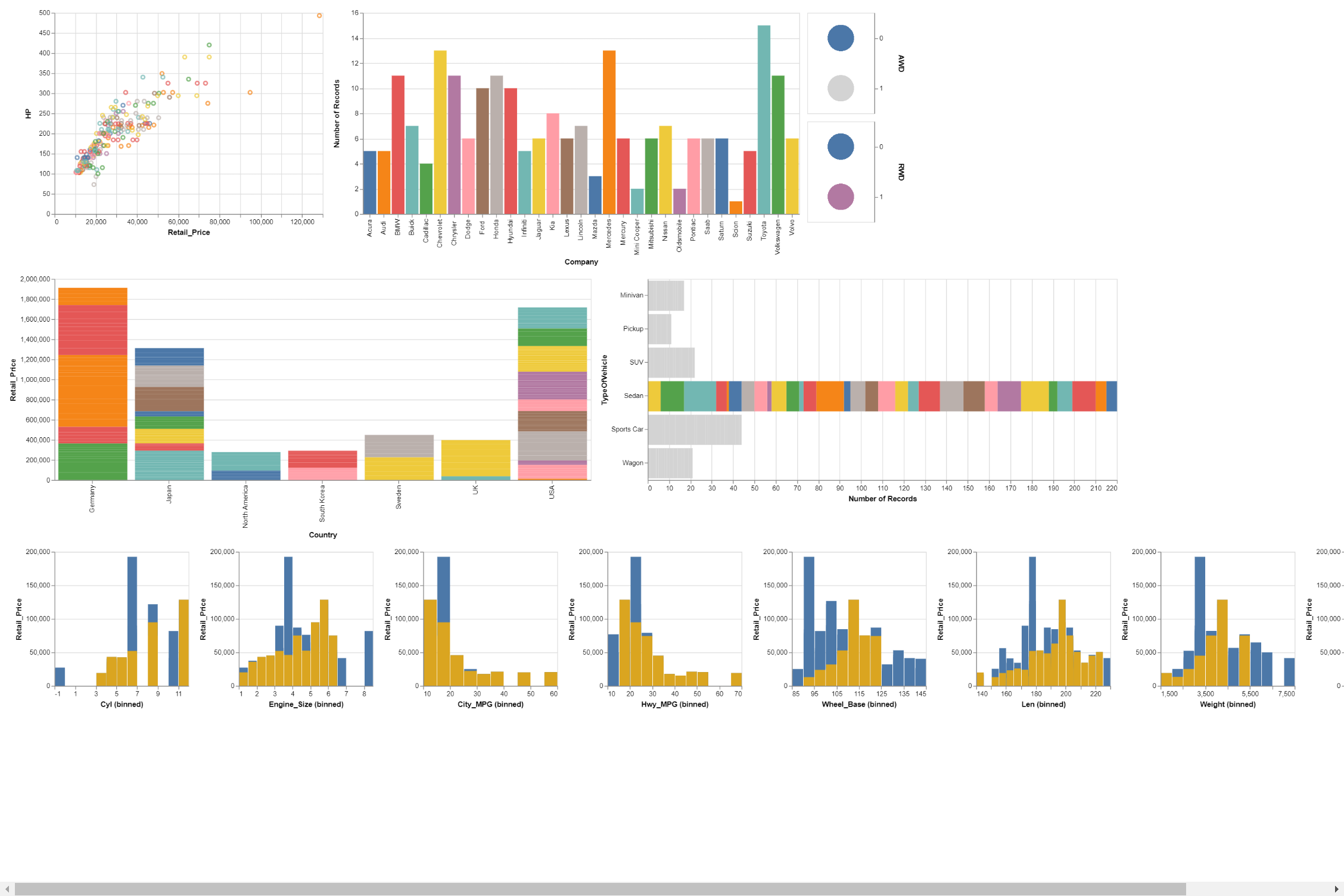


**Actions and analysis on the dashboard:**

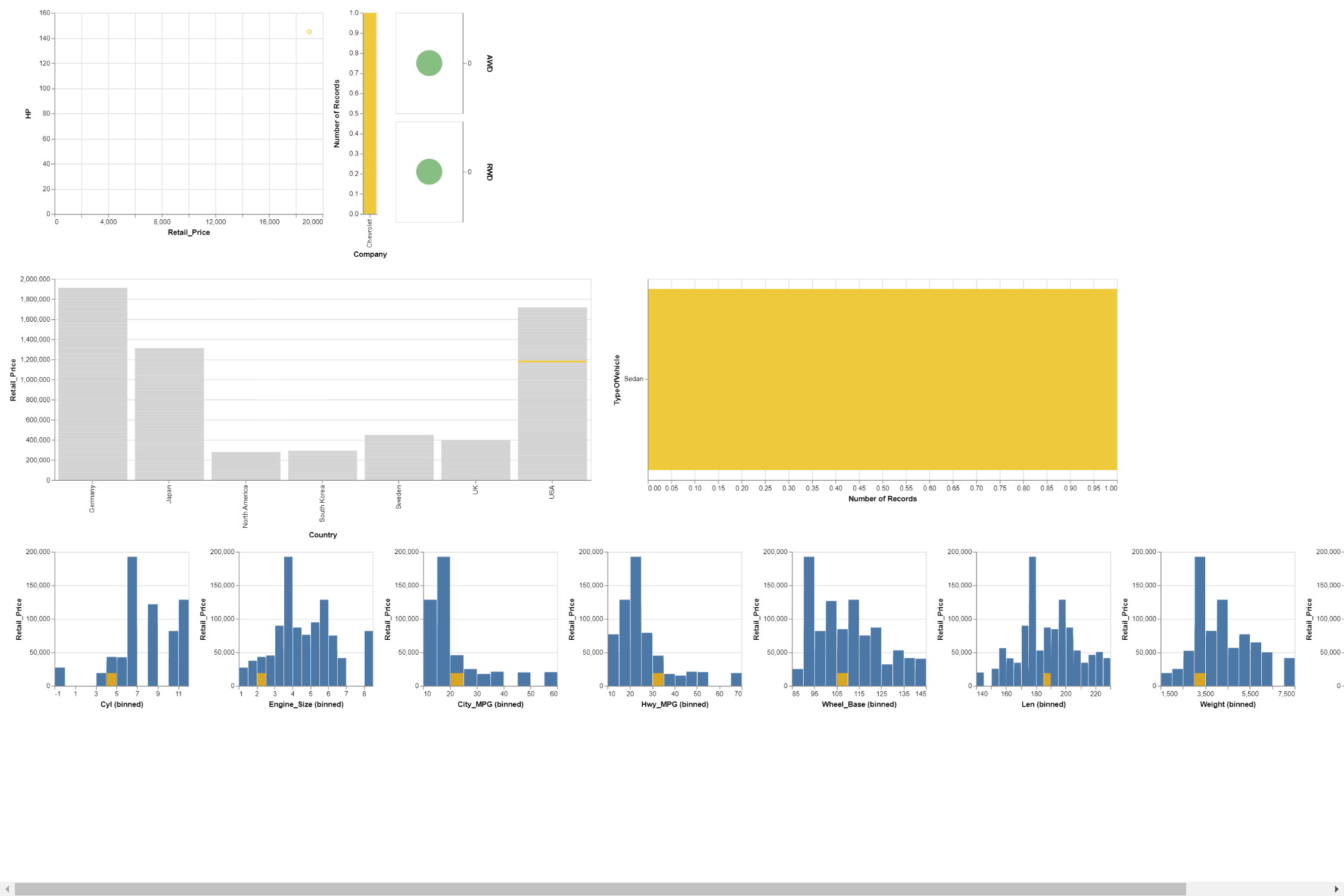
Selecting a single company: The dashboard highlights the graphs filtered for ex. Mercedes.



**On selecting a particular car type:**



**On filtering a particular model from country:**



Appendix: Source Code

**import** **altair** **as** **alt**

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**import** **io**

df2 = pd.read\_excel("https://raw.githubusercontent.com/pdhoot16/DataViz/master/04cars**%20d**ata.xls")

df2.head()

input\_dropdown = alt.binding\_select(options=list(set(df2.TypeOfVehicle)))

interval = alt.selection\_multi(fields=['Company'])

brush = alt.selection(type='interval', resolve='global')

carName = alt.selection\_single(fields=['VehicleName'])

carType = alt.selection\_single(fields=['TypeOfVehicle'])

brush\_tool = alt.selection(type='interval', encodings=['x'])

AWDselection = alt.selection\_multi(fields=['AWD'])

RWDselection = alt.selection\_multi(fields=['RWD'])

color1 = alt.condition(interval,

alt.Color('Company:N',legend=**None**),

alt.value('lightgray'))

color2 = alt.condition(brush,

alt.Color('Company:N',legend=**None**),

alt.value('lightgray'))

color3 = alt.condition(carName,

alt.Color('Company:N',legend=**None**),

alt.value('lightgray'))

color4 = alt.condition(carType,

alt.Color('Company:N', legend = **None**),

alt.value('lightgray'))

*#####################*

points = alt.Chart(df2).mark\_point().encode(

x='Retail\_Price',

y='HP',

tooltip=['VehicleName','HP','Retail\_Price'],

color=color2

*#alt.condition(interval, 'Company', alt.value('lightgray'))*

).add\_selection(

brush

).transform\_filter(

interval

).transform\_filter(

carType

).transform\_filter(

carName

).transform\_filter(

brush\_tool

).transform\_filter(

AWDselection

).transform\_filter(

RWDselection

)

*#####################*

histogram = alt.Chart(df2).mark\_bar().encode(

x='Company',

y='count()',

tooltip=['Company','count()'],

color=color1

).transform\_filter(

brush

).transform\_filter(

carName

).transform\_filter(

carType

).transform\_filter(

brush\_tool

).transform\_filter(

AWDselection

).transform\_filter(

RWDselection

).properties(

selection=interval

)

*#####################*

bar = alt.Chart(df2).mark\_bar().encode(

x="Country",

y="Retail\_Price",

color=color3,

tooltip=['VehicleName','Country','Retail\_Price']

).properties(

width = 800

).transform\_filter(

brush

).transform\_filter(

interval

).transform\_filter(

carType

).transform\_filter(

brush\_tool

).transform\_filter(

AWDselection

).transform\_filter(

RWDselection

).add\_selection(

carName

)

*#####################*

carTypeBar = alt.Chart(df2).mark\_bar().encode(

x=alt.X('count()'),

y='TypeOfVehicle',

tooltip=[alt.X('count()'),'Country','TypeOfVehicle'] ,

color=color4

).transform\_filter(

brush

).transform\_filter(

interval

).transform\_filter(

carName

).transform\_filter(

brush\_tool

).add\_selection(

carType

).transform\_filter(

AWDselection

).transform\_filter(

RWDselection

).properties(

width=700,

height=300

)

*#####################*

*# Define the base chart, with the common parts of the*

*# background and highlights*

base = alt.Chart().mark\_bar().encode(

x=alt.X(alt.repeat('column'), type='quantitative', bin=alt.Bin(maxbins=20)),

y='Retail\_Price',

tooltip=['Retail\_Price']

).properties(

width=200,

height=200

)

*# blue background with selection*

background = base.properties(selection=brush\_tool)

*# yellow highlights on the transformed data*

highlight = base.encode(

color=alt.value('goldenrod')

).transform\_filter(

brush\_tool

).transform\_filter(

brush

).transform\_filter(

interval

).transform\_filter(

carName

).transform\_filter(

carType

).transform\_filter(

AWDselection

).transform\_filter(

RWDselection

)

*# layer the two charts & repeat*

common\_bar = alt.layer(

background,

highlight,

data=df2

).repeat(column=["Cyl", "Engine\_Size", "City\_MPG","Hwy\_MPG","Wheel\_Base","Len","Weight","Width"])

*################*

awdPoint = alt.Chart(df2

).mark\_point(filled=**True**,size=1500

).encode(

y=alt.Y('AWD:O', axis=alt.Axis(orient='right', title='AWD')),

color=alt.condition(AWDselection,

'AWD:N',

alt.value('lightgray'),

legend=**None**),

tooltip=['AWD'],

).add\_selection(

AWDselection

).transform\_filter(

brush

).transform\_filter(

interval

).transform\_filter(

carType

).transform\_filter(

brush\_tool

).transform\_filter(

carName

).transform\_filter(

RWDselection

).properties(

width=100,

height=150

)

*################*

rwdPoint = alt.Chart(df2

).mark\_point(filled=**True**,size=1500

).encode(

y=alt.Y('RWD:O', axis=alt.Axis(orient='right', title='RWD')),

color=alt.condition(RWDselection,

'RWD:N',

alt.value('lightgray'),

legend=**None**),

tooltip=['RWD'],

).add\_selection(

RWDselection

).transform\_filter(

brush

).transform\_filter(

interval

).transform\_filter(

carType

).transform\_filter(

brush\_tool

).transform\_filter(

carName

).transform\_filter(

AWDselection

).properties(

width=100,

height=150

)

*################*

(points | histogram | (awdPoint & rwdPoint)) & (bar | carTypeBar) & common\_ba

R

References

Dataset - <https://www.idvbook.com/teaching-aid/data-sets/2004-cars-and-trucks-data/>

GitHub Link - <https://github.com/pdhoot16/DataViz>